

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Statistical properties and time-frequency analysis of temperature, salinity and turbidity measured by the MAREL Carnot station in the coastal waters of Boulogne-sur-Mer (France)

### Journal Item

#### How to cite:

Kbaier Ben Ismail, Dhouha; Lazure, Pascal and Puillat, Ingrid (2016). Statistical properties and time-frequency analysis of temperature, salinity and turbidity measured by the MAREL Carnot station in the coastal waters of Boulogne-sur-Mer (France). *Journal of Marine Systems*, 162 pp. 137–153.

For guidance on citations see [FAQs](#).

© 2016 Elsevier B.V. All rights reserved.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.jmarsys.2016.03.010>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

# Statistical properties and time-frequency analysis of temperature, salinity and turbidity measured by the MAREL Carnot station in the coastal waters of Boulogne-sur-Mer (France)

Ben Ismail Dhouha Kbaier Ben <sup>1,\*</sup>, Lazure Pascal <sup>2</sup>, Puillat Ingrid <sup>2</sup>

<sup>1</sup> IFREMER, REM/RDT/LSCM, France

<sup>2</sup> IFREMER, ODE/DYNECO/PHYSED, France

\* Corresponding author : Dhouha Kbaier Ben Ismail, email address : [dhouha.kbaier@ifremer.fr](mailto:dhouha.kbaier@ifremer.fr)

## Abstract :

In marine sciences, many fields display high variability over a large range of spatial and temporal scales, from seconds to thousands of years. The longer and longer recorded time series, with an increasing sampling frequency, in this field are often nonlinear, nonstationary, multiscale and noisy. Their analysis faces new challenges and thus requires the implementation of adequate and specific methods. The objective of this paper is to highlight time series analysis methods already applied in econometrics, signal processing, health, etc. to the environmental marine domain, assess advantages and inconvenients and compare classical techniques with more recent ones. Temperature, turbidity and salinity are important quantities for ecosystem studies. The authors here consider the fluctuations of sea level, salinity, turbidity and temperature recorded from the MAREL Carnot system of Boulogne-sur-Mer (France), which is a moored buoy equipped with physico-chemical measuring devices, working in continuous and autonomous conditions. In order to perform adequate statistical and spectral analyses, it is necessary to know the nature of the considered time series. For this purpose, the stationarity of the series and the occurrence of unit-root are addressed with the Augmented-Dickey Fuller tests. As an example, the harmonic analysis is not relevant for temperature, turbidity and salinity due to the nonstationary condition, except for the nearly stationary sea level datasets. In order to consider the dominant frequencies associated to the dynamics, the large number of data provided by the sensors should enable the estimation of Fourier spectral analysis. Different powerspectra show a complex variability and reveal an influence of environmental factors such as tides. However, the previous classical spectral analysis, namely the Blackman-Tukey method, requires not only linear and stationary data but also evenly-spaced data. Interpolating the time series introduces numerous artifacts to the data. The Lomb-Scargle algorithm is adapted to unevenly-spaced data and is used as an alternative. The limits of the method are also set out. It was found that beyond 50% of missing measures, few significant frequencies are detected, several seasonalities are no more visible, and even a whole range of high frequency disappears progressively. Furthermore, two time-frequency decomposition methods, namely wavelets and Hilbert-Huang Transformation (HHT), are applied for the analysis of the entire dataset. Using the Continuous Wavelet Transform (CWT), some properties of the time series are determined. Then, the inertial wave and several low-frequency tidal waves are identified by the

application of the Empirical Mode Decomposition (EMD). Finally, EMD based Time Dependent Intrinsic Correlation (TDIC) analysis is applied to consider the correlation between two nonstationary time series.

**Keywords :** Continuous wavelet transform, cross-correlation, empirical mode decomposition, Hilbert-Huang transform, stationarity, time dependent intrinsic correlation, time series, wavelets

## 1. Introduction

The birth of observatories contributes to improve the environmental knowledge. Indeed, long-term research infrastructures such as NEPTUNE Canada and in Europe: EMSO, FIXO3 and JERICO are providing longer and longer time series with high frequency records in open sea and coastal areas. The analysis of such time series requires the implementation of procedures and methods suitable to this increasing data flow towards an automatisation. Consequently, the objective of this paper is to explore adequate analysis methods, assess advantages and inconvenients and compare classical techniques with more recent ones. Coastal ecosystems, especially those of transitional waters, display high variability over a large range of spatial and temporal scales. To study their dynamics and estimate their variations at all scales, high frequency measurements are needed (Chang and Dickey, 2001; Chavez et al., 1997; Dickey, 1991; Dickey et al., 1993) together with appropriate analysis techniques. We present here the data recorded every twenty minutes by an autonomous monitoring device, settled in coastal waters of Boulogne-sur-Mer (eastern English Channel, France), recorded from 2005 to 2009. The fixed buoy station can record various biogeochemical parameters simultaneously. The time series of sea level, temperature, salinity and turbidity are considered. The fluctuations of the four quantities are important for ecosystem studies. The paper is structured in the seven following sections. After the description of the background and objectives in section 1, section 2 is dedicated to the presentation of the data. Then, stationarity issues are addressed in section 3. A preliminary analysis of stationarity focuses on the graphs and the correlograms. The conclusions from the direct observation of these are reinforced with application of the Augmented Dickey-Fuller (ADF) tests (Dickey and Fuller, 1979). Section 4 is devoted to a spectral analysis of the time series. First, a harmonic analysis is described. Secondly, the Blackman-Tukey spectral analysis is performed and comparisons between different periodograms are made. The problem of missing data is described and the Lomb-Scargle algorithm (Lomb, 1976; Scargle 1981) is used. This alternative does not require any interpolation of the data, in contrast to the Blackman-Tukey method. Furthermore, the limits of the Lomb-Scargle are illustrated. Continuous Wavelet Transform (CWT) and Hilbert- Huang Transformation (HHT) approaches are presented in section 5, while section 6 deals with an Empirical Mode Decomposition (EMD) based cross correlation method: Time Dependent Intrinsic Correlation (TDIC). Finally, a conclusion section (section 7) provides also possible developments of this study.

## Section 2: Presentation of the experimental database

The MAREL system (Automatic monitoring network for littoral environment, Ifremer, France, <http://www.ifremer.fr/marel>) is based on the deployment of moored buoys equipped with physico-chemical measuring devices, working in continuous and autonomous conditions (Berthome, 1994; Woerter, 1998). The measuring stations are equipped with high performance systems for seawater analysis and real time data transmission.

The database used in the present study comes from the MAREL Carnot system, located at the exit of the harbour of Boulogne-sur-Mer (France). This fixed point automated system records more than 15 physico-chemical parameters such as temperature, salinity, dissolved oxygen, pH, turbidity and water column level, with a high frequency resolution: an interval of 20 minutes. Some parts of these data set from the MAREL system have been analyzed by Prof. Schmitt and co-works: Huang and Schmitt (2014), Dur *et al.* (2007), Zongo and Schmitt (2011), etc. Their studies mainly focus on the scaling/scale invariance behavior.

In this paper, the dynamics of temperature, salinity and turbidity records, from January 2005 to the end of 2009, are investigated. The temperature sensor covers measurement from  $-5^{\circ}\text{C}$  up to  $+30^{\circ}\text{C}$ , with accuracies of  $0.1^{\circ}\text{C}$ . For the salinity, the conductivity sensor measures values between 0 and 70 mS/cm, with accuracies of 0.3 mS/cm. The turbidity sensor measures values between 0 and 4000 NTU, with a precision of about 10% of the value. Portions of the time series are shown in Figs. 1-3, for the three studied quantities. They seem to be intermittent and influenced by the tidal cycle, with also an important stochastic component (river flow, turbulence) since the Marel Carnot station is strongly influenced by the Liane river discharge.

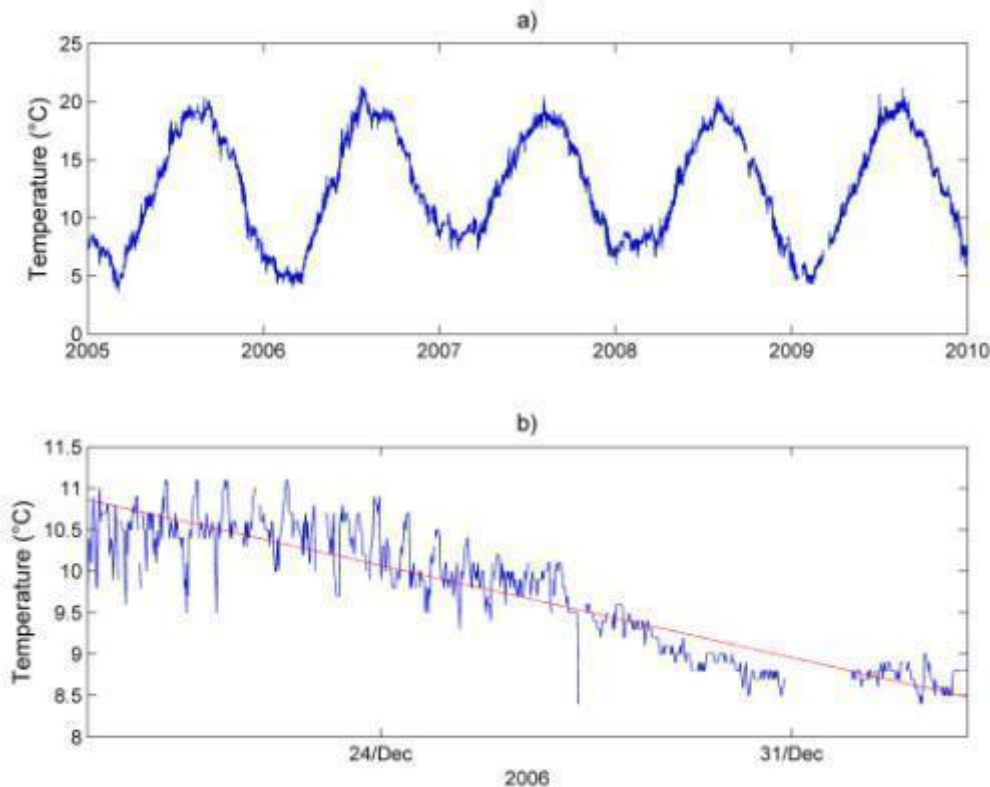


Figure 1. Temperature time series over 5 years (a), showing strong annual and seasonal cycles. A 15-day portion indicates a periodic component associated to the tide together with stochastic fluctuations (b). The red curve represents a downward trend.

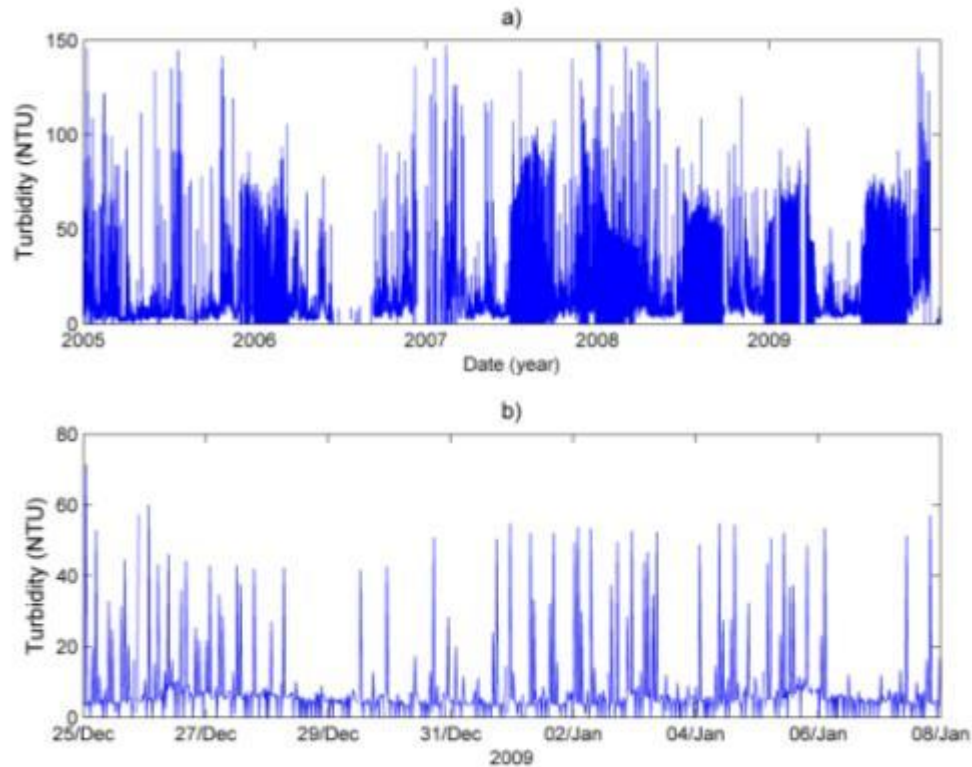


Figure 2. Turbidity time series over 5 years (a). A 15-day portion of the data sets indicates a strong periodic component associated to the tide together with stochastic fluctuations (b).

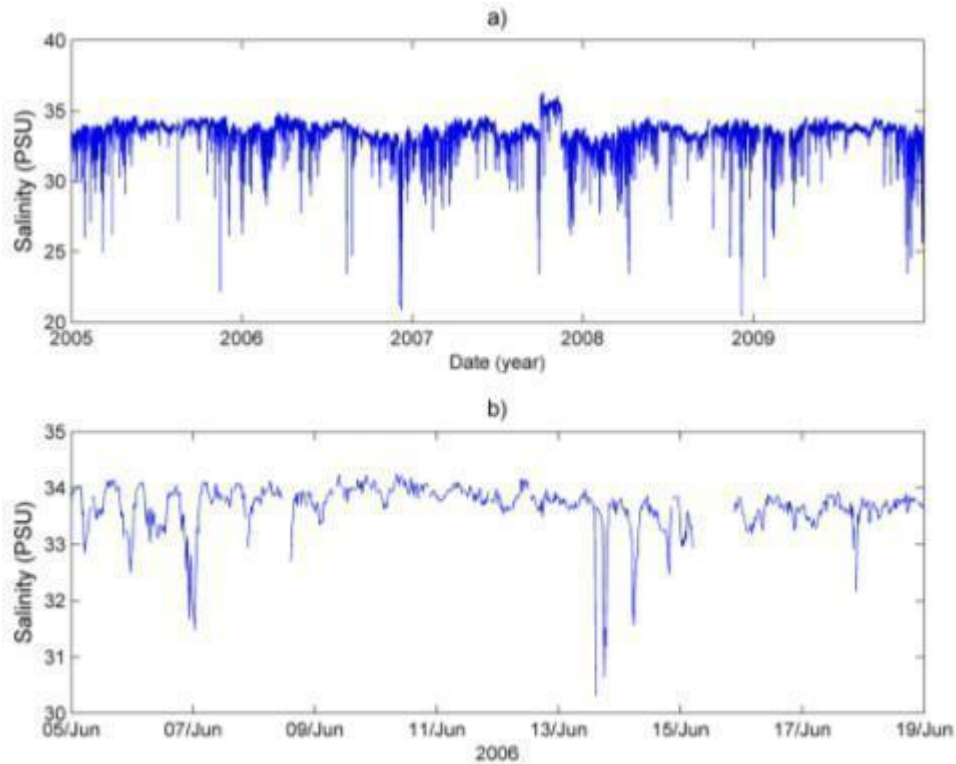


Figure 3. Salinity time series over 5 years (a). A 15-day portion of the data sets shows important and apparently irregular fluctuations (b).

Table 1 presents some statistical properties of the time series. The normalized third central moment is called the skewness and represents a measure of asymmetry of the tails of a distribution. The mode is the most common value, or the peak in the histogram. Fig. 4 displays different probability density functions (PDFs) of the data sets. We notice for the salinity that  $\text{mean} < \text{median} < \text{mode}$ , which corresponds to a negative skewness of  $-2.49$  and indicates that the distribution is spread out more to the left of the mean value (Panofsky and Brier, 1968). On the other hand, the distribution of the turbidity with a positive skewness has large tails that extend to the right. The distributions of temperature and sea level are almost symmetric. This observation is consistent with the corresponding values of skewness, since the skewness of the symmetric normal distribution is zero.

In Boulogne-Sur-Mer, the temperature of the sea remains almost the same during August and September as well as during February and March. Whereas the passage from the wintry temperatures to the summer temperatures is very quick and the offseason is very short. The net sheet of the heat due to the ocean-atmosphere exchanges is then a seasonal curve with a sinusoidal shape. Therefore, the histogram of temperature shows a bimodal pattern with two modes around  $9^{\circ}\text{C}$  and  $18^{\circ}\text{C}$ . Also, the sea level histogram shows a bimodal pattern where the two peaks correspond to slack tides, *i.e.*, high tide and low tide around 3 m and around 7 m respectively. The width of each peak is related to the Spring-Neap tidal cycle.

Before estimating the statistical models, it is important to analyze whether the series are stationary or not. Stationarity issues are addressed in the next section.

	Mean	Variance	Median	Mode	Skewness
Salinity ( $\mu\text{mol/L}$ )	33.43	0,79	33.56	33.75	-2.49
Turbidity (NTU)	12.33	203,61	7.70	0	2.73
Temperature ( $^{\circ}\text{C}$ )	12.65	20,86	18.30	12.70	-0.03
Sea level (m)	4.91	4.82	4.85	2.81	0.04

Table 1. Statistical properties of the time series.

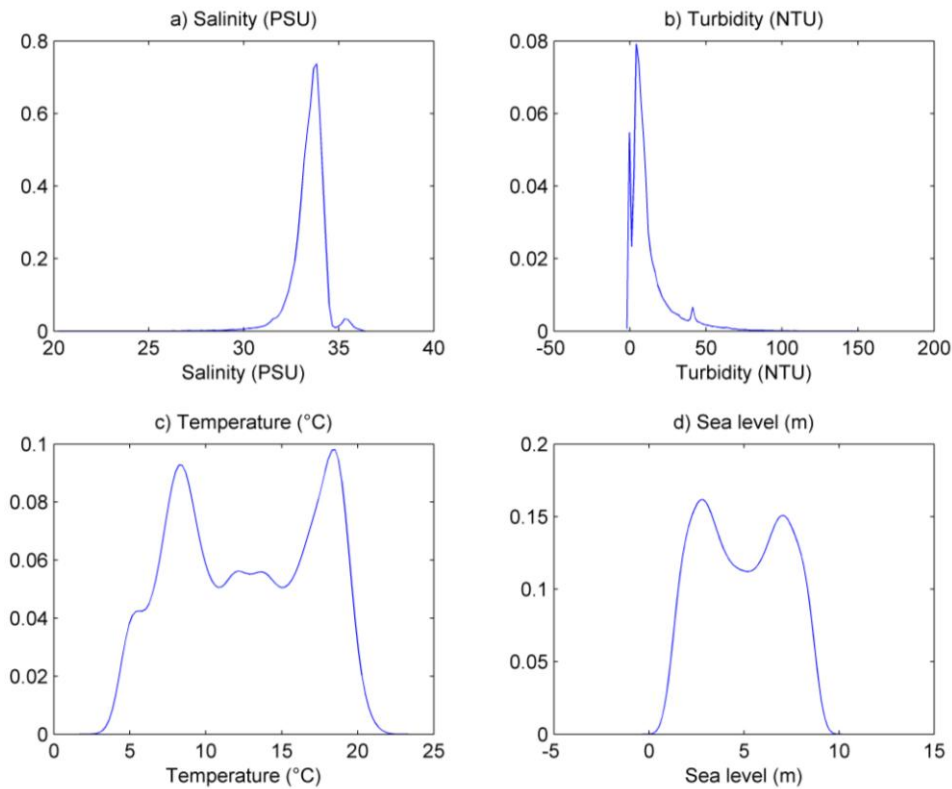


Figure 4. PDFs of a) salinity, b) turbidity, c) temperature and d) sea level.

### Section 3: Stationarity tests

The following definition of stationarity is taken from Challis and Kitney, 1991. “Stationarity is defined as a quality of a process in which the statistical parameters of the process do not change with time”. The “weak” stationary condition requires that the mean, variance and covariance structure of each variable do not change over time. It is important to analyze whether the series are stationary or not since it is not valid to use standard statistical tests in the negative case (Nelson and Plosser, 1982). Estimating the correlation between nonstationary variables can give spurious dependencies and conclusions about the development of temperature can be erroneous. If the dependent variable has a deterministic trend, a trend-adjusted series could be used instead. In case the variable has a unit-root, thus it has to be differenced. Further details about stationarity and its implications are available in Emery and Thompson (2014).

A preliminary analysis of stationarity is based on the graphs of the time series as well as the correlograms. Correlograms were used in early work such as the seminal



paper by Yule (1926), the monograph by Wold (1938) and the analysis of agricultural price series by Kendall (1943). As shown in Figs. 1-3, the graphs of temperature, turbidity and salinity are characterized by:

- A non stationarity in mean showed by a slight linear upward or downward trend. For instance, the red linear fitting plotted in Fig. 1 (b) shows how the temperature values are declining and represents a downward trend for the considered 15 day portion.
- An increase or decrease in variance by slices, characterizing a non stationarity in variance. For example, Fig. 5 displays the variance by slices for the temperature time series over the year 2007. The continuous and random increase and decrease in variance by slices illustrates the non stationarity of the time series. Here, the variability at time scale of the week is mainly induced by the variability of current (tides and wind induced circulation) rather than by the seasonal cycle of ocean-atmosphere fluxes.
- A regularly repeated phenomenon from the season to the tidal period (*i.e.*, from about one hour to six months), which lets suppose the presence of seasonality.

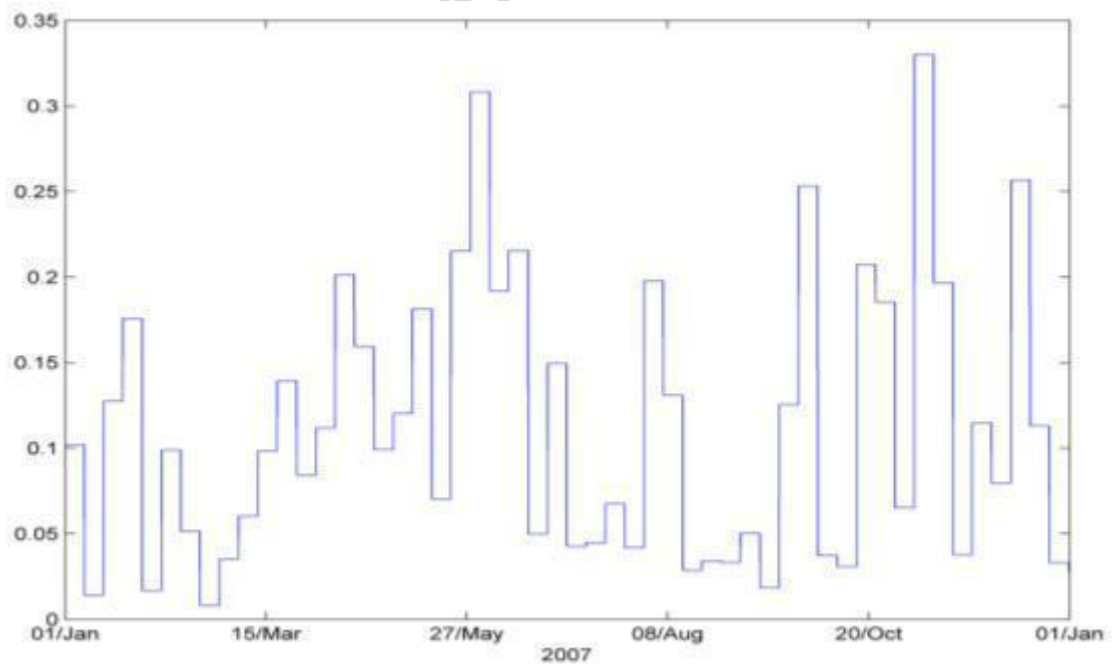


Figure 5. Variance by slices (window size of one week) for the temperature time series over the year 2007.

For the sea level plotted in Fig. 6, the time series is stationary since there is no systematic change in mean (no trend) and variance. Regarding the correlogram, Fig. 7 shows that it has few significant spikes at small lags (lag <5); and decreases drastically at 8 or 9 lags. Therefore, the correlogram confirms that the sea level series is stationary. Nevertheless, the correlogram of the turbidity in Fig. 8 is not declining and shows persisting significant values, indicating that the time series is nonstationary. Differencing is a transformation of the series to a new time series where the values are the differences between consecutive values. This procedure may be applied iteratively



more than once, giving rise to the "first differences", "second differences", etc. Differencing makes the time series stationary, as the correlogram has only a couple of significant spikes at lags 1 and 2. Similar results are obtained for salinity and temperature.

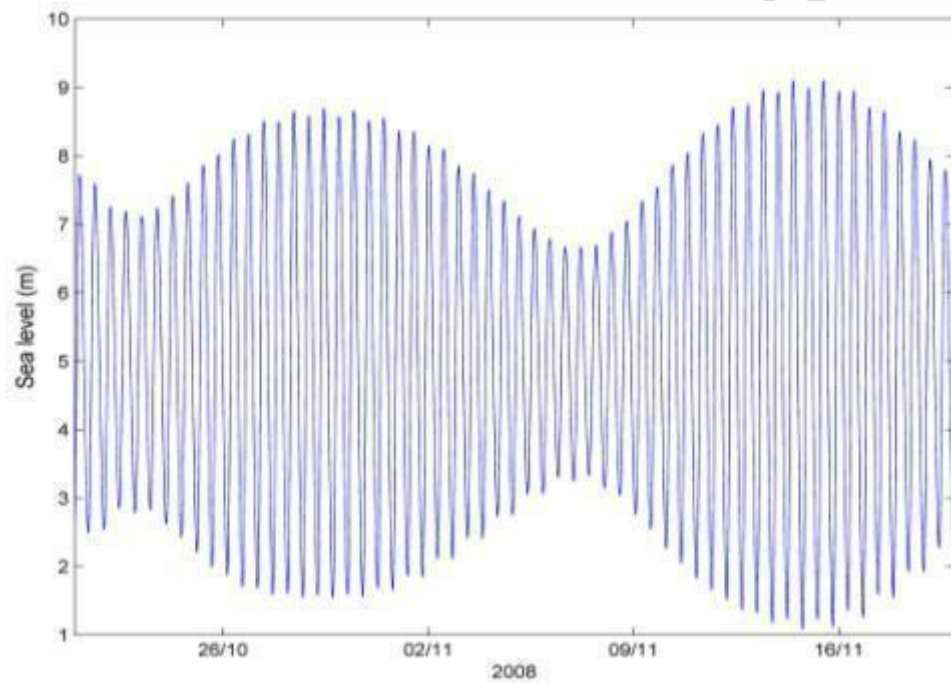


Figure 6. A 30-day portion of the sea level data sets indicates a strong periodic component associated to the tide.

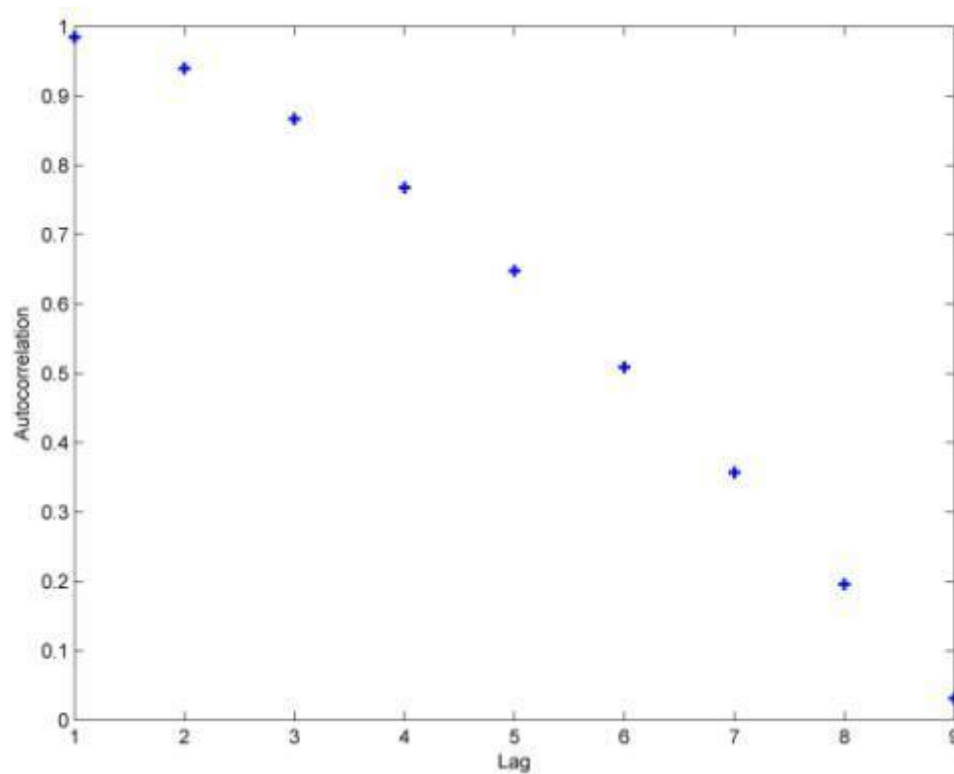


Figure 7. Correlogram of sea level data sets.

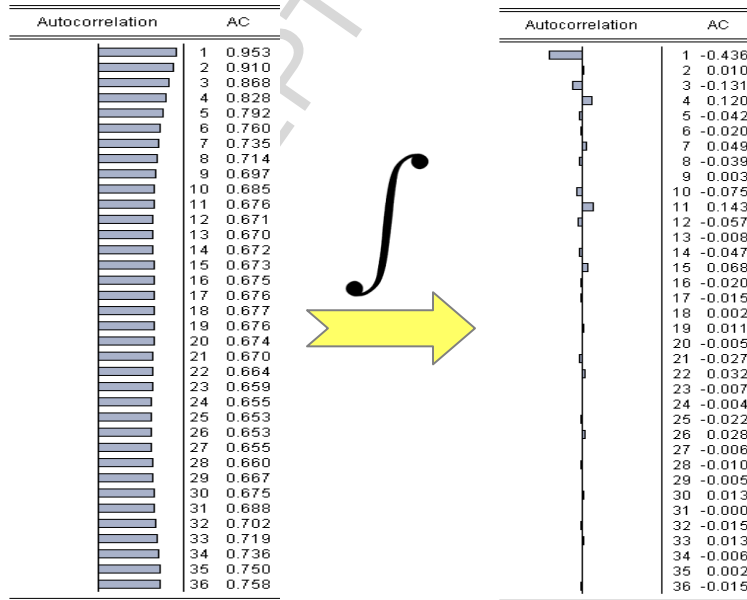


Figure 8. Correlogram of turbidity data sets before and after differencing.

After a preliminary inspection of the graphs, we move to a quantitative study of the stationarity. Indeed, using non-stationary time series data in environmental models may lead to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by

differencing. On the other hand, if the time series data analyzed exhibits a deterministic trend, the spurious results can be avoided by detrending. Furthermore, the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend. Sometimes it can be hard to tell the difference between a series that is trend-stationary and one that is difference-stationary. Therefore, the ADF unit root tests were used to get a more definitive answer. These are statistical hypothesis tests of stationarity that are designed for determining more objectively whether differencing is required. That is why we have applied Augmented Dickey-Fuller (ADF) tests (Dickey and Fuller, 1979) for testing of unit root and stationarity. Unit root corresponds to the application of the difference operator  $\Delta$  to induce stationarity. Given an observed time series  $(Y_1, Y_2, \dots, Y_N)$ , Dickey and Fuller consider three differential-form autoregressive equations to detect the presence of a unit root:

1.  $\Delta Y_t = \gamma Y_{t-1} + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \xi_t$
2.  $\Delta Y_t = \alpha + \gamma Y_{t-1} + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \xi_t$
3.  $\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^p \phi_j \Delta Y_{t-j} + \xi_t$

where  $t$  is the time index,  $\alpha$  is an intercept constant called a drift,  $\beta$  is the coefficient on a time trend,  $\gamma$  is the coefficient presenting process root, i.e. the focus of testing,  $p$  is the lag order of the first-differences autoregressive process and  $\xi_t$  is an independent identically distributes residual term. The difference between the three equations concerns the presence of the deterministic elements  $\alpha$  (a drift term) and  $\beta t$  (a linear time trend). The focus of testing is whether the coefficient  $\gamma$  equals to zero, what means that the original  $Y_1, Y_2, \dots, Y_N$  process has a unit root; hence, the null hypothesis of  $\gamma = 0$  (random walk process) is tested against the alternative hypothesis  $\gamma < 0$  of stationarity. If the original series needs differencing, then the coefficient  $\gamma$  should be approximately zero. If  $Y_t$  is already stationary, then  $\gamma < 0$ .

More detailed, the null and alternative hypotheses corresponding to the models above are:

- ( $h_1$ )  $H_0: \gamma = 0 \Leftrightarrow Y_t$  is random walk  
 $H_1: \gamma < 0 \Leftrightarrow Y_t$  is stationary process
- ( $h_2$ )  $H_0: \{\gamma = 0, \alpha \neq 0\} \Leftrightarrow Y_t$  is random walk around a drift  
 $H_1: \{\gamma < 0, \alpha \neq 0\} \Leftrightarrow Y_t$  is level stationary process
- ( $h_3$ )  $H_0: \{\gamma = 0, \beta \neq 0\} \Leftrightarrow Y_t$  is random walk around a trend  
 $H_1: \{\gamma < 0, \beta \neq 0\} \Leftrightarrow Y_t$  is trend stationary process

An important practical issue for the implementation of the ADF test is the specification of the lag length  $p$ . If  $p$  is too small then the remaining serial correlation

in the errors will bias the test. If  $p$  is too large then the power of the test will suffer. The results of the Monte Carlo experiments suggest that it is better to error on the side of including too many lags. Ng and Perron (1995) suggest the following data dependent lag length selection procedure that results in stable size of the test and minimal power loss.

- First, set an upper bound  $p_{\max}$  for  $p$ .
- Next, estimate the ADF test regression with  $p = p_{\max}$ .
- If the absolute value of the t-statistic for testing the significance of the last lagged difference is greater than 1.6 (see Ng and Perron, 1995) then set  $p = p_{\max}$  and perform the unit root test. Otherwise, reduce the lag length by one and repeat the process.
- A common rule of thumb for determining  $p_{\max}$ , suggested by Schwert

$$(1989), \text{ is } p_{\max} = \left\lceil 12 \times \left( \frac{N}{100} \right)^{1/4} \right\rceil \text{ where } [x] \text{ denotes the integer part of } x.$$

To estimate the significance of the coefficients in focus, the modified t (Student)-statistic (known as Dickey-Fuller statistic) is computed and compared with the relevant critical value (MacKinnon, 1996): if the test statistic is less than the critical value then the null hypothesis is rejected. Each version of the test has its own critical value which depends on the size of the sample.

The ADF tests results for salinity are presented in Fig. 9. The test of constant and a deterministic trend occurring results in the computed absolute ADF statistic (3.12), which is smaller than the tabulated critical value (11.62) at 90 % level of confidence. Therefore, the hypothesis that the time series is nonstationary and has a trend is accepted. The hypothesis of unit root with a drift is also accepted since the computed absolute ADF statistic (2.56) is smaller than the tabulated critical value (11.55) at 90 % level of confidence. However, the first ADF test without a constant is rejected since the test values are greater (in absolute value) than the critical value (0.36) for different confidence levels.

Similar results are obtained for temperature and turbidity data: all the critical values (MacKinnon, 1996) are the same and only the test values are different. Thus, the unit-root tests exclude that the series are pure random walk processes (first model). Nevertheless, these series are random walk processes with a drift (second model). Moreover, they are nonstationary and have a deterministic trend (third model). In the next section, we perform power spectral analysis of these data.

## Section 4: Power spectral analysis

### 4.1 Harmonic analysis

Harmonic analysis takes advantage of the fact that one knows *a priori* all the frequencies at which tidal energy will be found in a data time series. Most of these frequencies are astronomically caused but many are due to the nonlinear hydrodynamic effects of shallow water. Indeed, the shallower the water depth is the more the tidal wavelength will shorten and the faster the tidal characteristics of a waterway will change with horizontal (geographic) distance.

In classical harmonic analysis, the tidal signal is modeled as the sum of a finite set of sinusoids at specific frequencies related to astronomical parameters. The purpose of the tide analysis is to determine the *amplitude* and *phase* (the so-called *tidal harmonic constants*) of the individual cosine waves, each of which represents a *tidal constituent* identified by its *period* in mean solar hours. The tidal response is modelled as:

$$x(t) = \alpha + \beta t + \sum_{k=1}^N A_k \cos(\sigma_k t) + B_k \sin(\sigma_k t)$$

where  $N$  constituents are used. Each constituent has a frequency  $\sigma_k$  which is known from the development of the potential, and complex amplitudes ( $A_k$  and  $B_k$ ) which are not known. A possible offset and (optional) linear drift is handled by the first two terms  $\alpha$  and  $\beta$ .

One widely-used program for performing harmonic analysis of oceanic tides is the set of programs in MATLAB created by Pawlowicz et al. (2002) to perform classical harmonic analysis for periods of about 1 year or shorter. They have rewritten the FORTRAN package created by Mike Foreman (1977), incorporating many useful extras. Confidence intervals for the amplitudes and phases of the tidal constituents are calculated from the estimated variance. Confidence intervals were computed from the average of power spectral density within frequency bands centered on 0; 1; 2; ... cycles/day,  $\pm 0.1$  cycles/day ( $M_0$  and  $M_1$ ),  $\pm 0.2$  cycles/day ( $M_2$  to  $M_5$ ),  $\pm 0.21$  cycles/day ( $M_6$ ), 0.26 - 0.29 cycles/day ( $M_7$ ), and 0.3 - 0.5 cycles/day ( $M_8$ ) (Pawlowicz et al., 2002).

We have performed harmonic analysis of the data sets presented in section 2 “Presentation of the experimental database”. For the sea level, the harmonic analysis is able to predict 99.7% of the variance of the original signal (see Fig. 9). The highest tidal amplitudes and the corresponding waves can be found in Table 2. Thus, the semidiurnal wave  $M_2$  is the most energetic one. For the salinity, the harmonic analysis is able to predict 18.4% of the variance of the original signal. This is due to freshwater contents (received from the rivers) and a seasonal cycle with higher variability linked to the regime of rivers and residual (subtidal) circulation. Regarding the turbidity, it contains tidal components, but also a biological signal such as the one induced by the plankton in the spring, high-frequency effects linked to storms and a cycle of tidal streams (suspension and advection) (Burchard and Baumert, 1998; Chant and Stoner, 2001; Goni et al., 2005). Such components make the time series non stationary. This is the reason why the harmonic analysis is even worse: only 8.3% of the original signal’s variance is predicted. Hence, the harmonic analysis is not relevant except for the sea level. An adequate spectral analysis is introduced hereafter and allows comparing different periodograms.

Amplitude	Tidal wave	Type
2.8431	Principal lunar $M_2$	Semidiurnal
0.9429	Principal solar $S_2$	Semidiurnal
0.5160	Larger elliptical lunar $N_2$	Semidiurnal
0.3478	Declinational lunar-solar $K_2$	Semidiurnal
0.3160	$M_4 = 2M_2$	Compound

Table 2. Important tidal constituents obtained from the harmonic analysis of sea level.

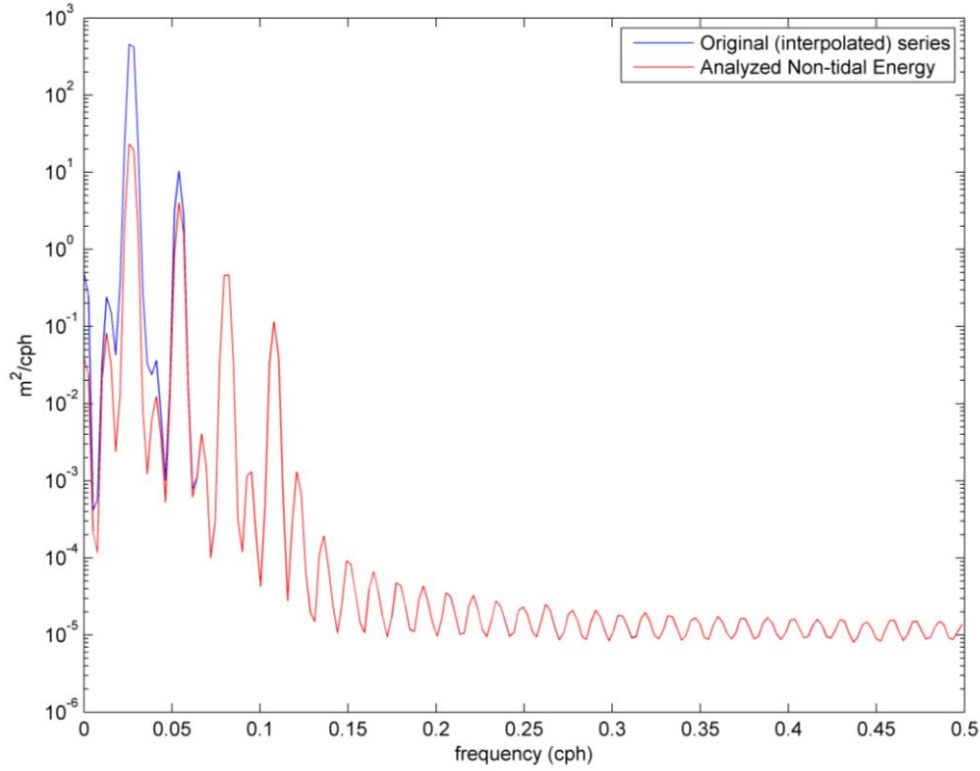


Figure 9. Use of `t_tide` toolbox for the harmonic analysis of sea level data: Spectral estimates before and after removal of tidal energy.

#### 4.2 Blackman-Tukey spectral analysis

The most popular method to compute powerspectra in earth sciences is the method introduced by Blackman and Tukey (1958). This method performs autospectral analysis in three steps: calculation of the autocorrelation sequence, windowing and finally computation of the discrete Fourier transform. For the modified periodogram, several window shapes are available such as Bartlett (triangular), Hamming (cosinusoidal) and Hanning (slightly different cosinusoidal). An enhancement was made by the application of the Welch's method (1967). First, the time series is divided into overlapping segments. After the data is split up, the overlapping segments are windowed, like for the modified periodogram. Then, the Welch method includes computing the powerspectrum for each segment and finally averaging the powerspectra. To be more specific, we divide an  $N$  point data sequence  $x(n)$ ,  $1 \leq n \leq N$ , into  $L$  segments of  $M$  samples each. The Welch periodogram is an  $L$  modified or windowed periodogram and can be defined as

$$I_M^l(f) = \frac{1}{UM} \left| \sum_{n=0}^{M-1} x_M^l(n) w(n) e^{-j\omega nT} \right|^2 \quad 1 \ll l \ll L$$

where  $U$ , the energy in the window is

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n)$$

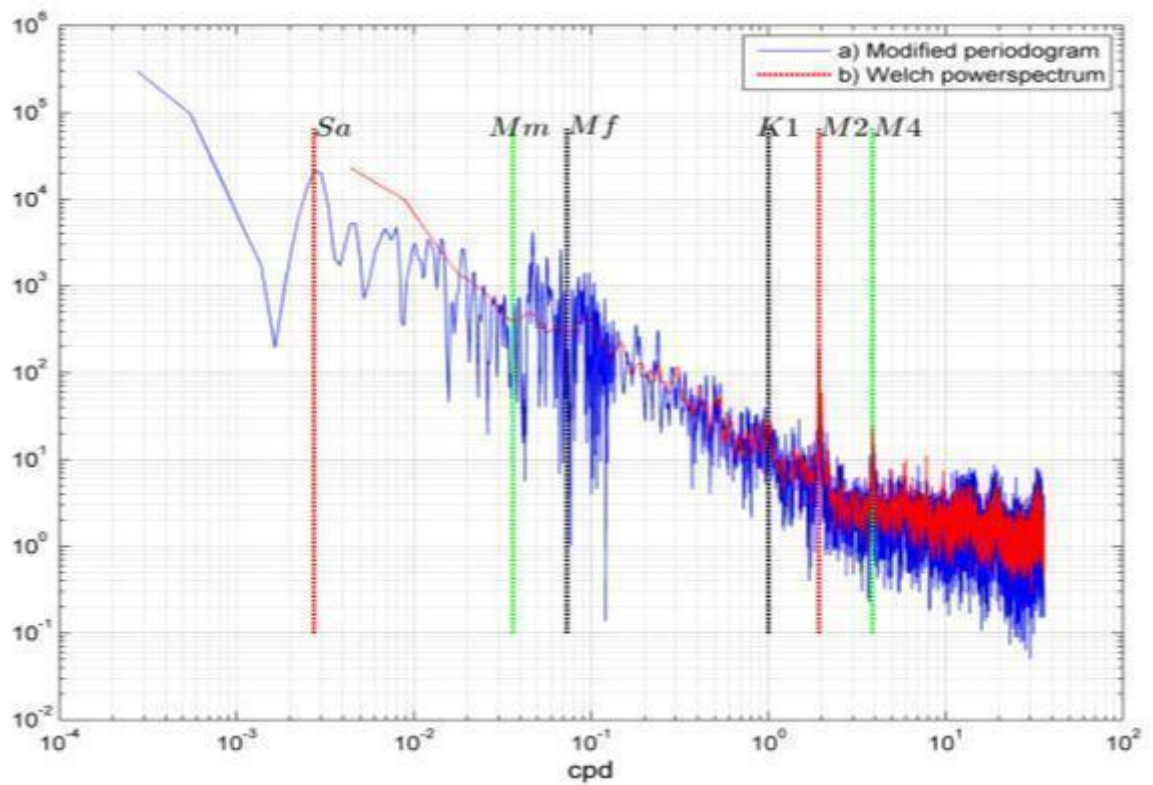


Figure 10. Two powerspectra of the turbidity data sets: modified periodogram with a Hamming window versus Welch powerspectrum.

The dynamics of turbidity is illustrated in Fig. 10 showing two powerspectra in log-log plot: the modified periodogram using a Hamming window and the Welch powerspectrum. Both reveal a very complex dynamics, with a turbulent-like behaviour at high frequencies, with a regime following roughly a power-law. This analysis is coherent with the power spectrum of turbidity obtained by Schmitt et al. (2008). For low frequencies, deterministic forcings are visible, associated with several energetic spikes such as the diurnal tidal cycle  $K_1$ , the semidiurnal cycle  $M_2$  and the corresponding propagation of  $M_4 = 2M_2$ . The diurnal period  $K_1$  (23,93 h) is very close to the day duration and the corresponding spike may hide biological effects on turbidity like the nyctemeral vertical migration of zooplankton. Compared to the modified periodogram, the Welch method improves the signal-to-noise ratio of the spectrum versus a loss of resolution. For example, there are not any spikes in the Welch powerspectrum for the monthly ( $Mm$ ) and semi monthly ( $Mf$ ) lunar tides due to an important loss of resolution at these low frequencies. Even more, the annual cycle ( $Sa$ ) is not detected at all. Indeed, for the Welch power spectrum, the raw data has been divided into several segment, the lowest frequency that could be resolved thus depends on the length of the segment.

In earth sciences, time series are often unevenly spaced. Missing values represent a common characteristic property of autonomous monitoring data bases due to routine maintenance, inaccessibility, vandalism, removal of biofouling, and failure of the measuring devices. The Blackman-Tukey method, however, requires evenly-spaced data. Therefore, we have interpolated the turbidity time series with 13% of missing data



in order to generate the powespectra in Fig. 10, as done in some studies (Ibanez and Conversi, 2002; Paparella, 2005). Nevertheless, interpolation introduces numerous artifacts to the data, both in the time and the frequency domain. For this reason, an alternative method of time-series analysis has become increasingly popular in earth sciences, the Lomb-Scargle algorithm (Lomb, 1976; Scargle 1981, 1982, 1989; Press et al. 1992; Schulz and Stattegger, 1998).

### 4.3 Lomb-Scargle powerspectrum

The Lomb-Scargle approach was done in the field of astronomy, corresponding to a framework which can be quite different from time series in coastal oceanography. In contrast to the Blackman-Tukey method, the Lomb-Scargle algorithm evaluates the data of the time series only at times  $t_i$  that are actually measured. Detailed discussions of the Lomb-Scargle method are given in Scargle (1989) and Press et al. (1992). An excellent summary of the method and a TURBO PASCAL program to compute the normalized Lomb-Scargle powerspectrum of paleoclimatic data has been published by Schulz and Stattegger (1998). A comfortable MATLAB algorithm *lombscargle* to compute the Lomb-Scargle periodogram has been published by Brett Shoelson (The MathWorks Inc.).

We suppose that a physical variable  $X$  is measured at a set of time  $t_i$ . Then, the Lomb-Scargle peridogram  $P_X$  as a function of angular frequency  $\omega = 2\pi f > 0$  is given by

$$P_X(\omega) = \frac{1}{2} \left( \frac{[\sum_j X_j \cos \omega(t_j - \tau)]^2}{\sum_j \cos^2 \omega(t_j - \tau)} + \frac{[\sum_j X_j \sin \omega(t_j - \tau)]^2}{\sum_j \sin^2 \omega(t_j - \tau)} \right)$$

where  $\tau$  is defined by

$$\tan(2\omega\tau) = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}$$

In the next, we consider the salinity time series over the year 2005, where 18.76% of the values are missing. Instead of filling the discontinuities with interpolated data, the spectral analysis is performed using the Lomb-Scargle algorithm (see Fig. 11). Similarly to the turbidity, a turbulent-like stochastic behaviour is visible at high frequencies, whereas the dynamics at low frequencies are dominated by deterministic forcings. Several energetic spikes are observed for the original time series of salinity. The first period corresponds roughly to the year ( $Sa$ ), then the monthly ( $Mm$ ) and semi monthly ( $Msf$ ) lunar tides. Furthermore, the diurnal and semidiurnal cycles are identified as for the previous analysis of turbidity. The peak between  $Mm$  and  $Msf$  seems to be more pronounced ( $f = 5 \times 10^{-2}$  days). It corresponds to a period of 20 days which is longer than an atmospheric disturbance. Thus, it could be related to the river discharges.

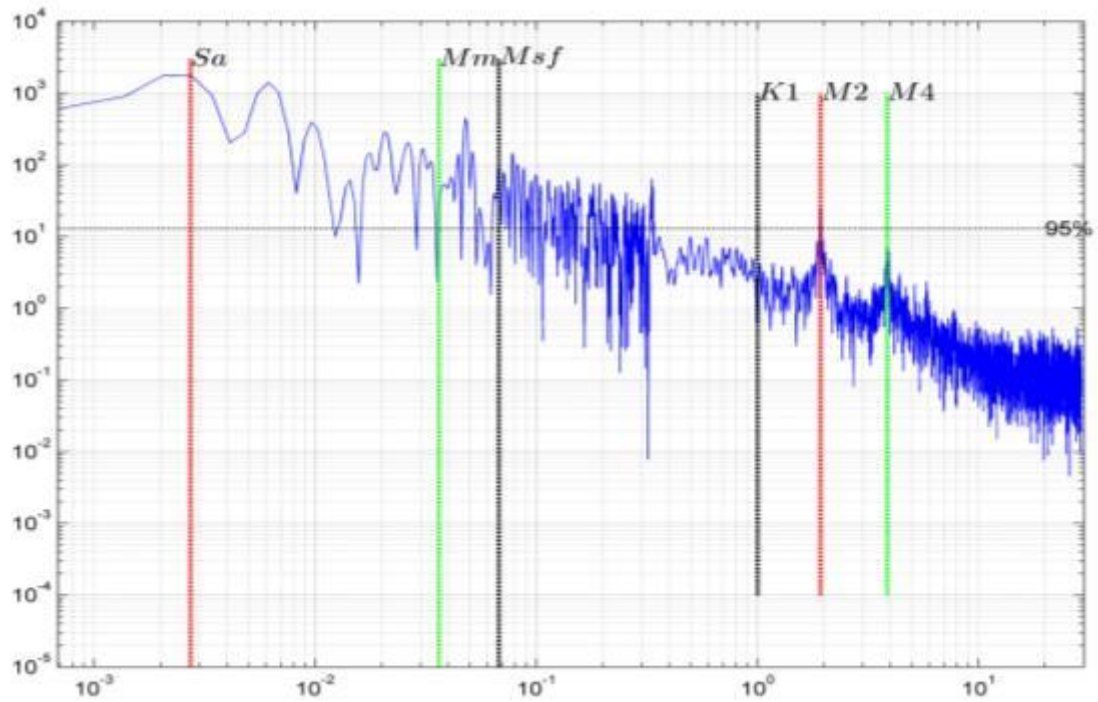


Figure 11. Lomb-Scargle powerspectrum of the salinity data over the year 2005.

The method is adapted for unevenly spaced data, however, one might ask the following questions. What are the limits? If the time series has a high rate of missing values, would the Lomb-Scargle powerspectrum be significant? Would the method always detect the most energetic frequencies? Or would it suffer above a certain threshold? To answer these questions, we have removed intentionally and randomly other data from the original time series as follows:

1. Case 1: Remove randomly one month of measures, representing 8.22% of additional missing values. Thus, the rate of missing values is increased from 18.76% to nearly 27%.
2. Case 2: Remove randomly 3 months of measures, representing 24.65% of additional missing values. Thus, the rate of missing values is increased from 18.76% to 43.41%.
3. Case 3: Remove randomly 6 months of measures, representing 49.31% of additional missing values. Thus, the rate of missing values is increased from 18.76% to nearly 68%.
4. Case 4: Remove randomly 90% of the data.

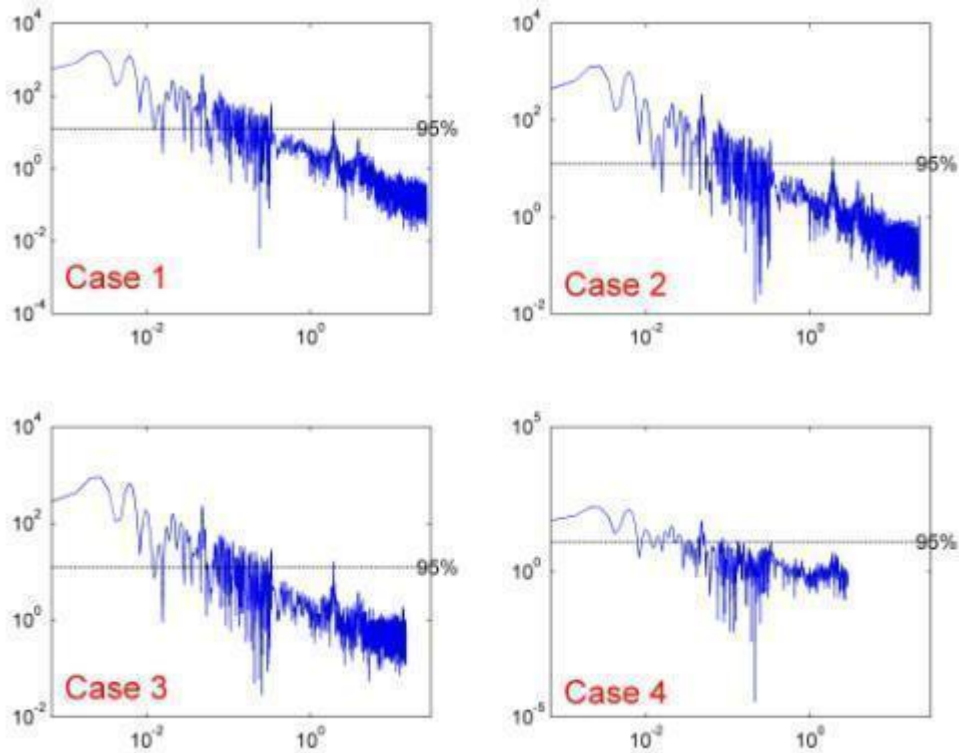


Figure 12. Comparing different Lomb-Scargle powerspectra for the salinity while increasing the rate of missing data.

Fig. 12 displays the corresponding Lomb-Scargle powerspectra. The more data are discarded, the less energetic spikes are observed in the powerspectrum. Table 3 provides the number of significant frequencies above the 95% confidence level, dropping down from 131 for the original data of salinity to only 14 when 90% of the data are randomly eliminated. Several seasonalities are no more visible in the spectra and even a whole range of high frequency disappears progressively. Beyond 3 months of discarded data, the Lomb-Scargle algorithm really suffers from a high rate that approaches 50% of missing measures.

Time series	≠ significant frequencies
Original data	131
Case 1	90
Case 2	79
Case 3	53
Case 4	14

Table 3. Number of significant frequencies above the 95% confidence level detected by the Lomb-Scargle method.

Although the Fourier transform is a traditional spectral analysis method, it is predicated on *a priori* selection of basis functions. The Fourier transform is suitable for analyzing periodic signals like ocean tides or Earth tides. The signal is compared with harmonic oscillations of different frequencies. The Fourier Transform gives information about the frequency content of the signal without any time localization. Thus, time-

varying amplitudes and/or periods can hardly be detected. Consequently, the Fourier analysis is also constrained by the assumptions of stationarity and linearity. To accommodate the variety of data generated by nonlinear and nonstationary processes, Huang *et al.* (1998, 1999) developed a new adaptive data analysis method designated by the NASA as the Hilbert-Huang Transform (HHT) and introduced hereafter. Unlike the Fourier transform, which transforms information from the time domain into the frequency domain, the HHT does not lose temporal information after transformation, *i.e.* energy–frequency information is maintained in the time domain.

## Section 5: Hilbert Huang Transform

While the natural physical processes are mostly nonlinear and nonstationary, the available data analysis methods are either for linear but nonstationary processes such as wavelet analysis and the Wagner-Ville distribution, or nonlinear but stationary processes (see, for example, Tong 1990; Kantz and Schreiber 1997; Diks 1999). The development of the HHT was motivated by the need to describe nonlinear distorted waves in detail, along with the variations of these signals that naturally occur in nonstationary processes (Huang *et al.*, 1998). The HHT consists of the following steps:

- sifting, that is, empirical adoption of the Principal Component Analysis (PCA) for multi-components in the signal to rearrange the signal in terms of local bases that are nearly orthogonal each other;
- physically based construction of instantaneous frequency whose concept is applicable to nonstationary and nonlinear signals;
- complexification of the signal via the Hilbert Transform to characterize the signal in terms of the modulated amplitude and the associated instantaneous frequencies that appear to represent both interwaves and intrawaves;
- reconstruction of the signal and the Hilbert spectrum (energy-frequency) and the multi-component frequency-time relations.

The HHT technique for analyzing data consists of two components: a decomposition algorithm called Empirical Mode Decomposition (EMD) and a spectral analysis tool called Hilbert spectral analysis.

### 5.1 Empirical Mode Decomposition

The key part of the approach HHT is that any complicated dataset can be decomposed with the EMD method into a finite and small number of Intrinsic Mode Functions (IMFs), which represent different scales of the original time series and physically meaningful modes. An IMF is defined as a function having the same number of extrema and zero-crossings. It has also symmetric upper and lower envelopes defined by the local maxima and minima respectively.

The IMFs represent different scales of the original time series and physically meaningful modes. Indeed, the EMD divides the time series into a series of modes, but unlike a Fourier transform where each component has a constant frequency and a cosine or sine shape, the EMD allows each mode to have a time-dependent frequency and amplitude. In contrast to almost all the previous methods, the EMD works directly in

temporal space rather than in the corresponding frequency space. With the definition of the IMFs, one can then decompose any function  $x(t)$  through a sifting process:

- First identify all the local extrema (minima) and then connect all the local maxima (minima) by a cubic spline line to form the upper (lower) envelope. The envelopes should encompass all the data between them.
- Calculate  $m_1$ , the running mean of the envelope.
- Treat the difference  $x(t) - m_1 = h_1$  as the data in the next iteration and repeat the above sifting processing until  $h_1$  becomes an IMF.
  - $h_1 - m_{1(1)} = h_{1(1)}$ .
  - After  $k$  times, the first IMF is derived:  $h_{1(k-1)} - m_{1(k)} = h_{1(k)} = c_1$ . The term  $c_1$  contains the highest frequency component of the original signal.
  - Separate  $c_1$  from the rest of the data:  $r_1 = x(t) - c_1$  and carry out the same process described above.
- Repeat the procedure until  $r_n$  becomes a constant, a monotonic function or a function with only one maximum and one minimum from which no more IMF can be extracted:
  - $r_1 - c_2 = r_2$
  - ...
  - $r_{n-1} - c_n = r_n$
- By summing all the equations, we can get a decomposition with  $n$  IMFs and a

$$x(t) = \sum_{j=1}^n c_j(t) + r_n(t)$$

residue  $r_n$  as follows:

Due to a dyadic filter bank property of the EMD algorithm (Flandrin et al., 2004; Huang et al., 2008; Wu and Huang, 2004), usually in practice, the number of IMFs modes is less than  $\log_2(N)$ , where  $N$  is the length of the data set. Note that the sifting process serves two purposes: (1) to eliminate background waves on which the IMF is riding and (2) to make the wave profile more symmetric. Unlike Fourier based methodologies such as Fourier analysis and wavelet transform, the EMD method does not define a basis a priori (Flandrin and Gonçalves, 2004; Huang et al., 1998, 1999) and the IMFs obtained by sifting processes constitute an adaptive basis. Therefore, EMD is very suitable for nonstationary and nonlinear time series analysis (Huang et al., 1998, 1999). One way to express the nonstationarity is to find instantaneous frequency and instantaneous amplitude. This was the reason why Hilbert spectrum analysis was included as a part of HHT.

HHT has already been applied in marine sciences. For example, Dätig and Schlurmann (2004) applied HHT to show excellent correspondence between simulated and recorded nonlinear waves. Schmitt *et al.* (2009) applied the HHT method to characterize the scale invariance of velocity fluctuations in the surf zone. The EMD scheme was used in studying sea level rise (Ezer et al., 2013). Yin *et al.* (2014) also applied the method and identified three kinds of low-frequency waves using some observations in the coastal water of the East China Sea.

In this paper, the EMD algorithm has been applied to the Marel Carnot data sets acquired from 2005 to 2009 and already presented in section 2 (Presentation of the experimental database). The analysis for the salinity records is presented below.

## 5.2 EMD of the salinity time series

After EMD decomposition of the salinity data sets, 19 IMFs modes are obtained plus the residual, as shown in Fig. 14 and Fig 15. The time scale is increasing with the number of the IMF mode, the first IMF thus corresponding to the highest frequency. The residual from EMD algorithm has been recognized as the trend of the given data (Moghtaderi et al., 2011; Wu et al., 2007).

From EMD, several tidal waves are identified through the IMFs: semidiurnal ( $IMF_6$ , 12 hours such as M2), fourth-degree diurnal ( $IMF_4$ , 6 hours such as M4), spring and neap tides, semimonthly ( $IMF_{13}$ ), monthly ( $IMF_{14}$ ), seasonal ( $IMF_{15}$ ), semiannual ( $IMF_{16}$ ) and annual waves ( $IMF_{17}$ ). The  $IMF_7$  may represent the inertial wave with a

theoretical period of  $\frac{1}{f_{inertial}} = \frac{1}{2 \times \sin(\varphi^\circ)} (days) = \frac{24}{2 \times \sin(\varphi^\circ)} (hours)$  (Monin, 1990).

Here, the inertial wave is  $\frac{24}{2 \times \sin(51^\circ)} = 15.44$  hours. The period of  $IMF_7$  is close to the inertial period but the bottom friction induced by the strong tidal currents is likely to dissipate these inertial oscillations and this period may be due the non linear interaction between tidal and high frequency atmospheric forcing.

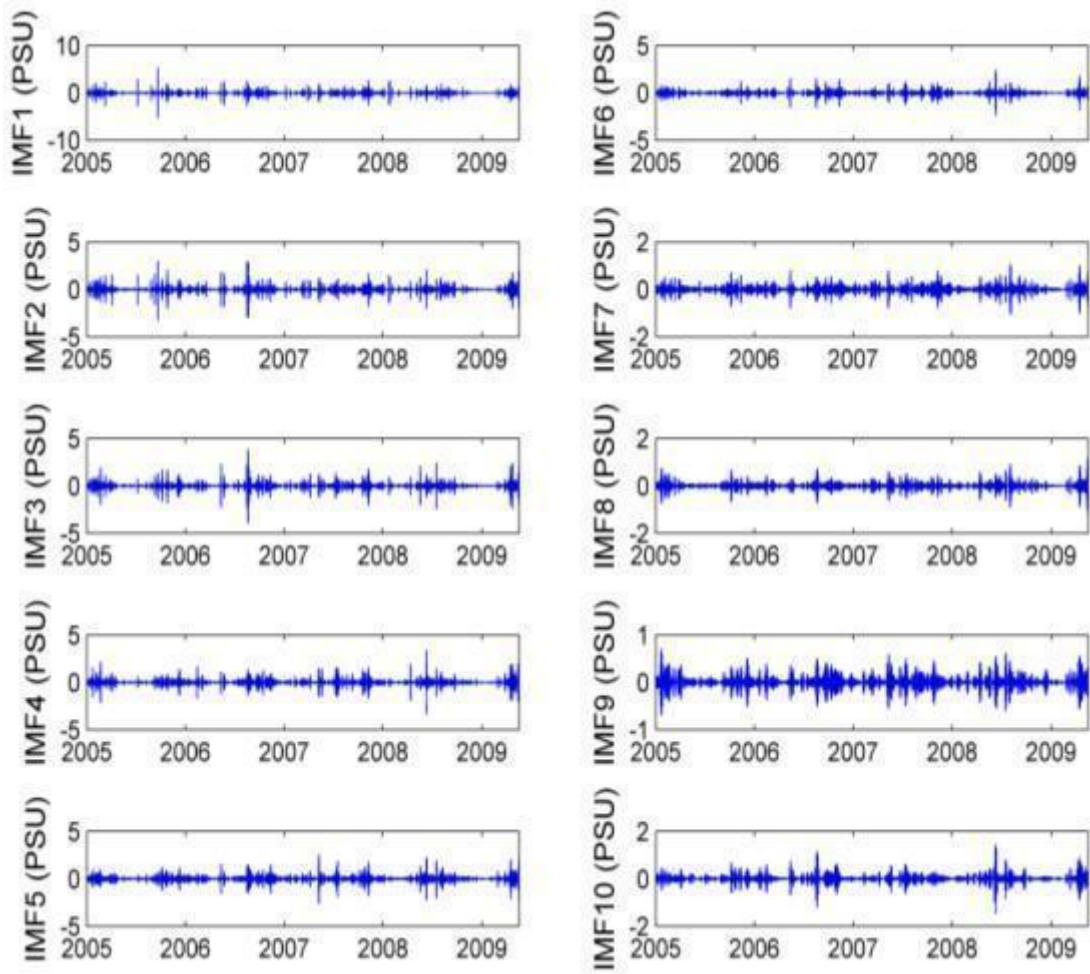


Figure 13. The first-tenth IMF modes for salinity. The time scale is increasing with the mode index.



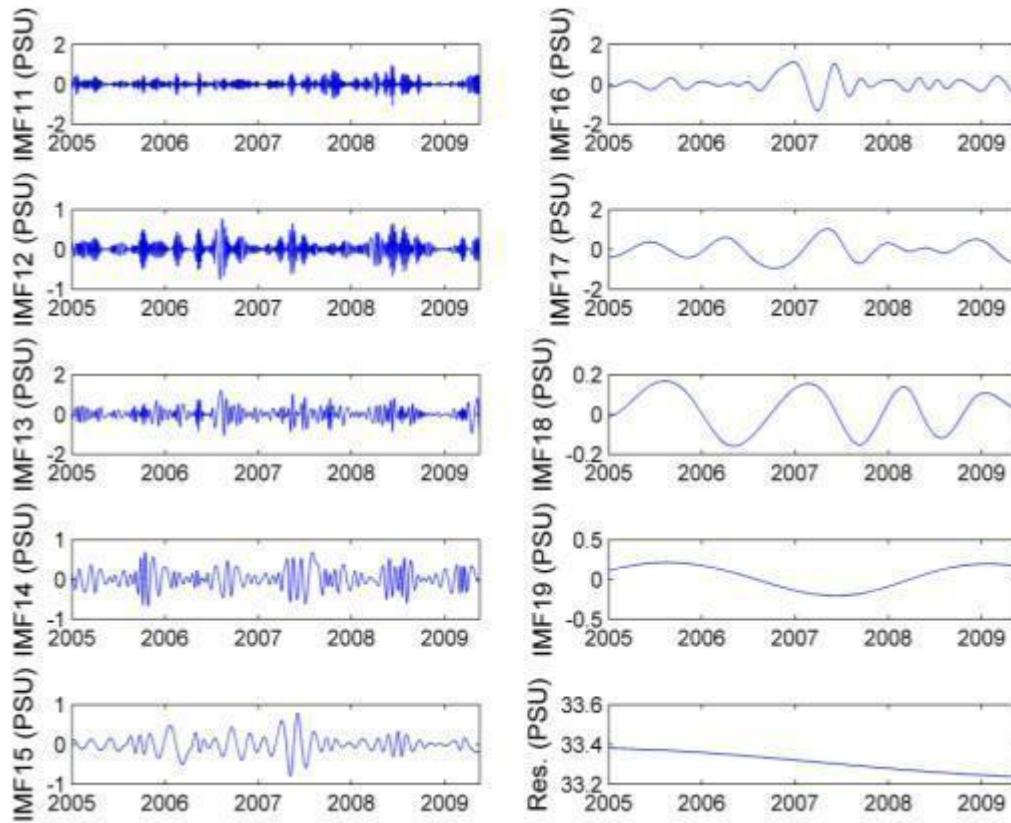


Figure 14. The last-nineteenth IMF modes together with the trend term for salinity.

Table 4 summarizes the mean period of the IMF modes and some tidal waves obtained after EMD decomposition of the salinity. Furthermore, the contribution of each IMF to the total energy is measured by the variance. Table 4 gives also the contribution of each IMF to the total energy (measured by the variance), percentage of total and cumulative percentage contribution to the sum of the variances. For the high frequencies, the first-seventh IMFs account for nearly 25% of the total energy. Thus, the main part of energy is not carried by the very high frequency components but by intermediate and low ones, especially the annual wave  $IMF_{17}$ . It can be seen that the biggest contribution comes from four most energetic IMFs: the annual wave ( $IMF_{17}$ ), the semiannual wave ( $IMF_{16}$ ), the semimonthly wave  $IMF_{13}$  and the monthly wave  $IMF_{14}$ . These four IMFs account for over 52 % of the total energy.

While the Fourier expansion would require tens of modes to represent the whole data, the EMD method decomposes the time series into only 19 IMFs plus the residual. When all the IMFs are added back successively, we notice that all the energy is recovered, as shown in all the cases in Huang et al. (1998, 1999).

IMFs	Mean period (days)	Tidal waves	Variance	% of total	Cumulative %
IMF1	0,06 = 1,43 hours		0.0292	3.04	3.04
IMF2	0,11 = 2,66 hours		0.0247	2.57	5.61
IMF3	0,18 = 4,44 hours		0.0359	3.74	9.35
IMF4	0,26 = 6,20 hours	M4	0.0423	4.40	13.76

IMF5	0,39 = 9,44 hours		0.0508	5.29	19.05
IMF6	0,52 = 12,59 hours	Semidiurnal M2	0.0377	3.93	22.98
IMF7	0,76 = 18,16 hours	Inertial	0.0179	1.87	24.84
IMF8	1,30 = 31,16 hours		0.0179	1.86	26.71
IMF9	1,99 days		0.0155	1.61	28.31
IMF10	3,48 days		0.0316	3.29	31.61
IMF11	6,06 days		0.0367	3.83	35.43
IMF12	8,97 days		0.0316	3.29	38.73
IMF13	17,68 days	Semimonthly	0.0921	9.59	48.32
IMF14	31,97 days	Monthly	0.0562	5.85	54.17
IMF15	63,35 days	Season (3 months)	0.0457	4.76	58.93
IMF16	170,11 days	Semiannual (6 months)	0.1553	16.18	75.11
IMF17	318,76 days	Annual	0.2049	21.34	96.45
IMF18	433,07 days		0.0102	1.06	97.51
IMF19	1011,05 days	~18 months	0.0218	2.27	99.78
Sum					100.00

Table 4. Contribution of each IMF to the total variance for  $Temp_2$  time series

To analyze the variability, both HHT and Continuous Wavelet Transform (CWT) can be used. Comparisons of HHT and wavelets are investigated in the next subsection.

### 5.3 Continuous Wavelet Transformation (CWT)

The wavelet transform is widely used as a time-frequency analysis technique to deal with nonstationary signals. The choice of the mother wavelet is usually dependent on the type of data to deal with (Torrence and Compo, 1998). HHT on the other hand does not require any convolution of the signal with a predefined basis function or mother wavelet. The process of decomposition is data-driven.

The CWT is a consecutive pass-band filter through time series, recognized as a very successful approach for the analysis of time series (Addison, 2010; Avdakovic et al., 2013; Torrence and Compo, 1998). However, similar to the Fourier transform and in contrast to the EMD algorithm, the wavelet transform requires evenly-spaced data. Therefore, the data have to be interpolated.

Mathematically, a wavelet transformation decomposes a signal  $y(t)$  into some elementary functions  $\psi_{a,b}(t)$  derived from a mother wavelet  $\psi(t)$  by dilation and translation,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

where  $b$  denotes the position (translation) and  $a$  ( $> 0$ ) the scale (dilation) of the wavelet (Lau and Weng 1995). The wavelet transform of the signal  $y(t)$  about the mother wavelet  $\psi(t)$  is defined as the convolution integral

$$W(b, a) = \frac{1}{\sqrt{a}} \int \psi\left(\frac{t-b}{a}\right) y(t) dt.$$

There are many mother wavelets available in the literature. However, in geosciences the most popular wavelet is the Morlet wavelet, which is given by

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}$$

where  $\eta$  is the non-dimensional time and  $\omega_0$  is the wavenumber (Torrence and Compo 1998). The wavenumber is the number of oscillations within the wavelet itself.

In Fig. 16, the wavelet transform has been applied to the four time series. Many high frequencies are detected for temperature, salinity and turbidity records. More variations are visible for the turbidity time series, which confirms that the data sets are highly non-stationary compared to the other data. This analysis confirms also the previous results presented in 4.1, where the performed harmonic analysis allows to predicting only 8.3% of the original signal's variance. As explained before, this is due to biological contents induced by the plankton in the spring, high-frequency effects linked to storms and a cycle of tidal streams (suspension and advection).

However, if we look at the CWT of the sea level in Fig. 16, the graph shows horizontal clusters of peaks at 365 days and 0.5 day, which corresponds to a strong annual cycle and the semidiurnal wave. The powerspectrum also reveals other significant periods such as the diurnal wave (1 day), the semimonthly and monthly waves; although these are sometimes not very clear.

As the signal content is now known in the frequency domain, the focus will shift to the application of an EMD based Time Dependent Intrinsic Correlation (TDIC), in order to study the possible links between physical observations.

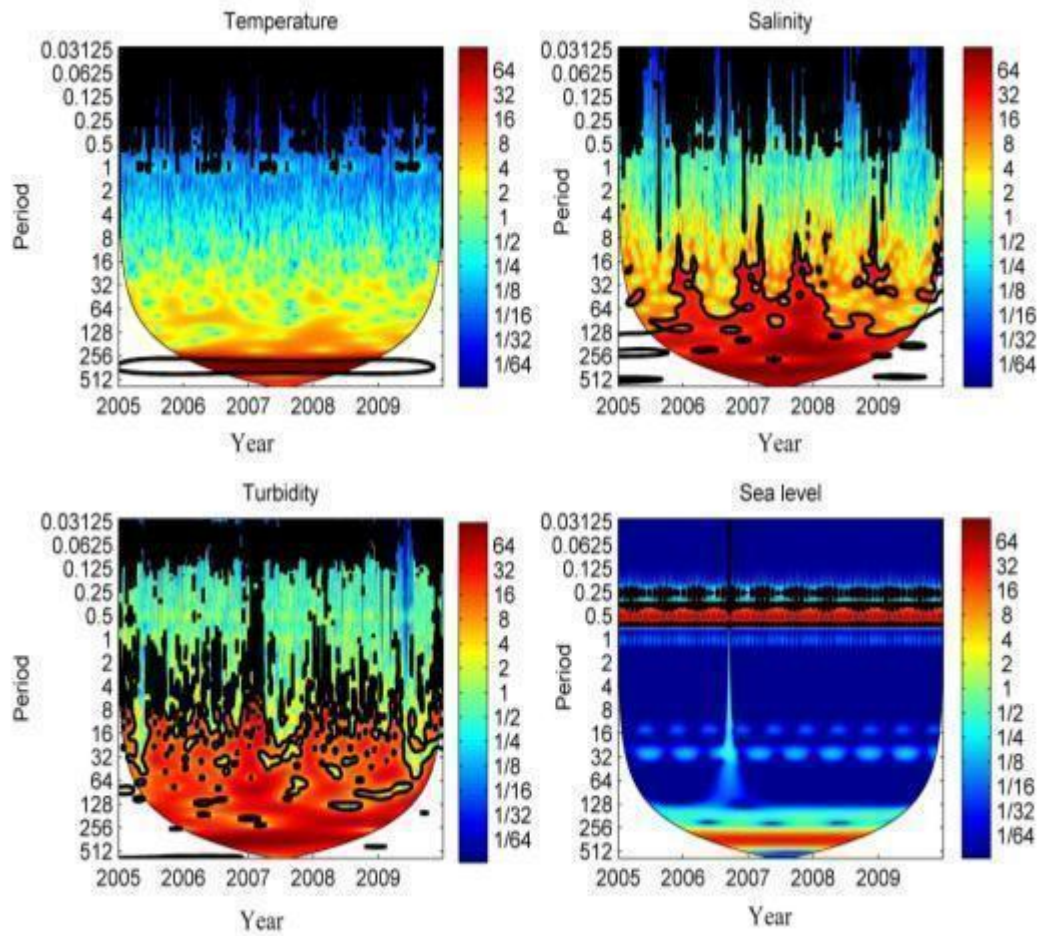


Figure 15. The Morlet wavelet spectrum of the data. The period in the y-axis is expressed in days.

## Section 6: Cross correlation methods

The classical global expression for the correlation (defined as the covariance of two variables divided by the product of the standard deviation of the two variables) assumes that the variables should be stationary and linear. Applied to nonstationary time series, the cross correlation information may be altered and distorted. The limitations of the correlation coefficient are also obvious: it is unable to provide local temporal information, and it cannot distinguish the main cycles from noise when measuring correlation. Many scientists tried to address the problem of nonsense correlations through different ways.

An alternative is to estimate the correlation coefficient by means of a time-dependent structure. For example, Papadimitriou et al. (2006) applied a sliding window to localize the correlation estimations. Rodo and Rodriguez-Aria (2006) developed the scale-dependent correlation technique. Although these methods detected the correlation between two nonstationary signals by computing the correlation coefficient in a local sliding window, the main problem is to determine the size of this window. Recently, Chen et al. (2010) introduced an approach based on EMD. They proposed to first decompose the nonlinear and nonstationary data into their IMFs, then use the instantaneous periods of the IMFs to determine an adaptive window and finally compute the time dependent intrinsic correlation coefficients. Huang and Schmitt (2014) used TDIC to analyze temperature and dissolved oxygen time series obtained from automatic measurements in a moored buoy station in coastal waters of Boulogne-Sur-Mer (France). In the following, we give an insight into the TDIC method.

### 6.1 Time Dependent Intrinsic Correlation (TDIC)

The correlation between two data sets is considered here. Suppose the two time series  $x_1(t)$  and  $x_2(t)$  can be represented in terms of their IMFs as  $x_i(t) = \sum_{j=1}^n h_j^i(t) + r_i(t)$  where  $h_j^i(t)$  is the  $j^{th}$  IMF of  $x_i(t)$  and  $r_i(t)$  are the residues. We find the mean period  $T_j^i(t)$  of each  $h_j^i(t)$  either by calculating the local extrema points and zero crossing points, i.e.,  $T_j^i(t) = 4 \times \frac{\text{data length}}{Nbr_{\max} + Nbr_{\min} + Nbr_0}$  (Huang and Schmitt, 2014) or by considering the

Fourier energy weighted mean frequency, i.e.,  $T_j^i(t) = \frac{\int |X_j^i(f)|^2 df}{\int f |X_j^i(f)|^2 df}$  where  $X_j^i(f)$  is the

Fourier power spectrum of each IMF mode. Then, at time  $t_{inst}$ , the sliding window is given by  $t_{win} = \left[ t_{inst} - a \frac{\max(T_j^1(t), T_j^2(t))}{2}, t_{inst} + a \frac{\max(T_j^1(t), T_j^2(t))}{2} \right]$ , where  $a$  is any positive number. This window is different from classical sliding windows: it is based on the maximum of two instantaneous periods  $\max(T_j^1(t), T_j^2(t))$  and thus it is adaptive.

### 6.2 Results for cross analyses

The focus will shift now to illustrate the cross correlations between temperature and sea level time series. We consider a sample of one year for these records between

24<sup>th</sup> May 2006 and 24<sup>th</sup> May 2007. First, the time series are not correlated since the global cross correlation coefficient is only 0.025.

Then, the EMD algorithm is applied to the data sets for the same time period between 24<sup>th</sup> May 2006 and 24<sup>th</sup> May 2007. For the temperature records, there are 12 IMFs modes with one residual, while the sea level is decomposed into 6 IMFs with one residual. Thus, the data are represented in a multiscale way (Flandrin and Gonçalves, 2004; Huang et al., 1998; Huang, 2009) and the IMFs can be used for multiscale correlation. For instance, let us consider the IMFs modes with a mean period of 12 hours (semidiurnal) :  $IMF_5$  for temperature and  $IMF_1$  for sea level data sets. The global correlation coefficient is 0.09 for the corresponding IMFs modes.

Fig. 17 displays the measured TDIC and shows rich patterns at small sliding window. We note a decorrelation of the TDIC and we observe less rich patterns with the increase of the window size (as the variable  $a$  increases). Although the global cross correlation is too small, the TDIC detects several periods of high correlation between the modes. To focus on these periods, Fig. 18 shows a zoom of the IMFs modes between 15<sup>th</sup> July and 15<sup>th</sup> August 2006. These are positively correlated with each other on some portions and negatively correlated on others; showing rich dynamics. The red circles point out some portions where the IMFs co-vary. This means that sometimes the temperature records are highly influenced by the tidal cycles. For example, the correlation coefficient between  $IMF_5$  and  $IMF_1$  for the time series of temperature and sea level respectively, is 0.59 from 15<sup>th</sup> to 20<sup>th</sup> July 2006, 0.61 from 4<sup>th</sup> to 7<sup>th</sup> August 2006, which is much higher compared to the global correlation coefficient between these IMFs, equal to only 0.09.

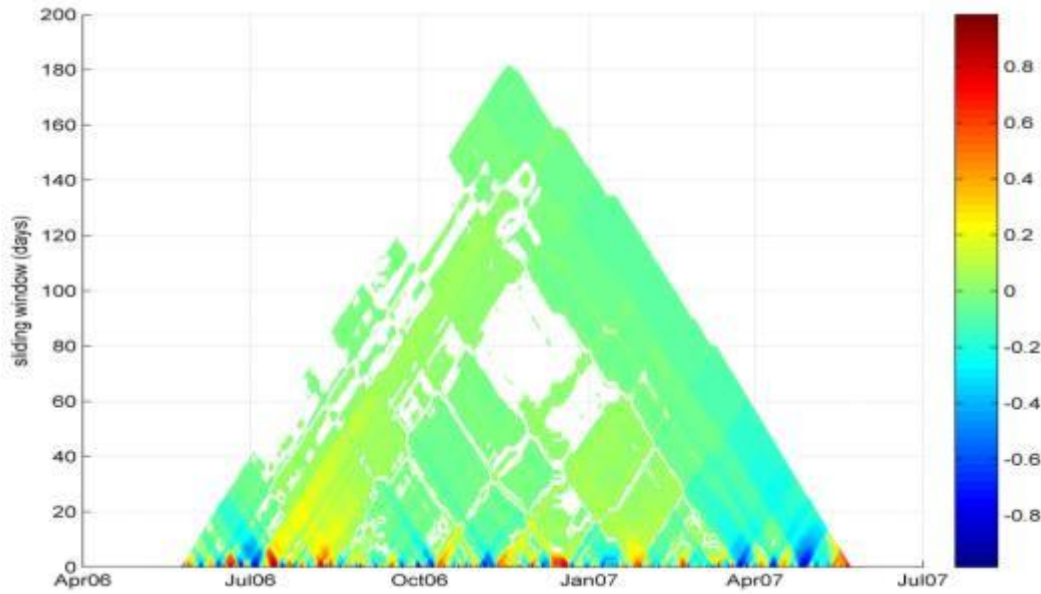


Figure 16. The measured TDIC for the 12-hour mean period obtained after EMD decomposition of temperature and sea level between 24<sup>th</sup> May 2006 and 24<sup>th</sup> May 2007.



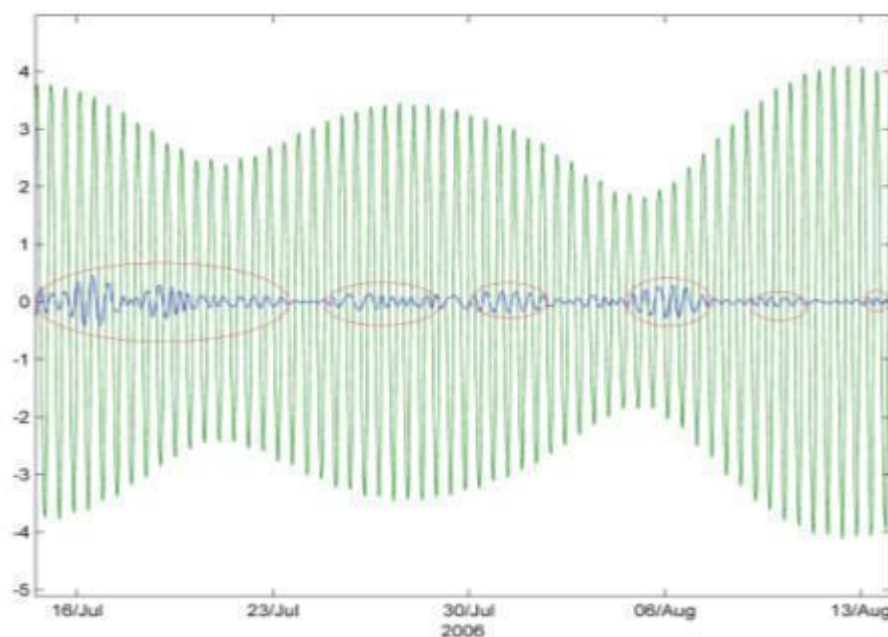


Figure 17. Zoom on the 12-hour cycle from EMD for temperature and sea level.

## Section 7: Conclusions and perspectives

In the era of “Big Data”, the time series are longer and longer. Time series analysis in marine sciences is becoming a challenge at the crossroads of marine observatories’ data. Indeed, the observatories allow more and more marine data to be safely archived. They contribute to understand and forecast the most crucial marine processes. Therefore, it is necessary to analyze this increasing data flow in an optimal way. In environmental sciences, adequate and specific methods are needed to analyze the complex time series which are nonstationary and reflecting nonlinear processes.

To explore these methods, we have analyzed here four time series recorded at high frequency in the coastal waters of Boulogne-sur-Mer, using the MAREL Carnot monitoring station: sea level, salinity, turbidity and temperature data sets. In order to perform adequate statistical and spectral analyses, it is necessary to know the nature of the considered time series. First, a preliminary analysis of stationarity, based on the graphs of the time series as well as the correlograms, is carried out. In contrast to the sea level time series which is stationary, the graphs of temperature, turbidity and salinity are characterized by a non stationarity in mean and variance. The series are also perturbed by effects such as the tide and the earth’s rotation. Regarding the corresponding correlograms, these show strong persistence since there are significant spikes at several lags, indicating that the time series are nonstationary. Then, the stationnarity of the series and the occurrence of unit-root are addressed with the ADF tests. For the salinity data, the time series is nonstationary. It has a unit-root with drift and a deterministic trend. Similar results are obtained for temperature and turbidity data.

After stationarity issues, an adequate power spectral analysis is performed. The harmonic analysis involves the search for the amplitude and phase of certain known frequencies, used for the calculation of the astronomical tide’s generator potential. Apart from the sea level where the harmonic analysis is able to predict 99.7% of the

original signal's variance, such analysis is not relevant for the other time series. Different spectra for the turbidity and salinity are investigated. They reveal a very complex dynamics, with a turbulent-like stochastic behaviour at high frequencies, whereas the dynamics at low frequencies are dominated by deterministic forcings. Compared to the modified periodogram, the Welch method improves the signal-to-noise ratio of the spectrum versus a loss of resolution. Since the Blackman-Tukey method requires evenly-spaced data, we have interpolated the time series. However, interpolation introduces numerous artifacts to the data, both in the time and the frequency domain. Thus, the Lomb-Scargle algorithm is adapted for unevenly-spaced data and has been used as an alternative method. To know the limits of this algorithm, we have investigated the time series of salinity over the year 2005 and we have intentionally increased the rate of missing values by discarding randomly data from the original time series. The Lomb-Scargle powerspectra show that, beyond 50% of missing measures, few significant frequencies are detected, several seasonalities are no more visible, and even a whole range of high frequency disappears progressively.

The authors also explored an other type of interesting direction of study since the Blackman-Tukey spectral analysis is not adapted to nonlinear and nonstationary data sets. An adaptive data analysis method, called EMD, was developed to accommodate the variety of data generated by nonlinear and nonstationary processes. In this paper, EMD was used to analyze the fluctuations and the correlations between different time series. For example, the salinity observations were decomposed with the EMD method into a finite and small number of IMFs (19 IMFs) which represent different scales of the original time series. These are also physically meaningful modes since a group of tidal waves were detected. Comparisons with CWT are addressed. Furthermore, the decomposition into modes helps also to estimate how correlations vary among scales. In fact, the authors apply a recent methodology, based on EMD and called TDIC, in order to display patterns of correlations at different scales for different IMFs modes. Finally, to illustrate this approach, cross correlations between temperature and sea level time series are considered. For the semidiurnal wave, the TDIC is displayed and allows to detecting several periods of high correlation.

We can mention here some perspectives opened by this work. Techniques of filtering will be investigated to study the high frequency or to concentrate on seasonalities. Classical Butterworth filters can be implemented or demerliac filters, already used by the "Service Hydrographique et Océanique de la Marine" (S.H.O.M) to calculate the daily mean sea level. Moreover, we have studied each parameter separately. A development of this approach will be to study the couplings between some parameters. Such coupling will be studied considering joint or conditional moments, joint probability density function, and joint spectral analysis.

More generally, time series analysis methods could be implemented in an automated way in the environmental databases in order to facilitate the data discovery as well as the investigation tools and techniques.

## Acknowledgments

The authors would like to thank the Région Bretagne for financial support of the post-doctoral fellowship (SAD MASTOC). They also thank the organizing committee of the MAREL Carnot symposium held in June 2014, Boulogne-Sur-Mer, France. Finally, we wish to thank the Editor, the Associate Editor, and the Referees for helpful comments.



## References

- ADDISON, P.S., “The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance”, Taylor & Francis, New York, 2010.
- AVDAKOVIC, S., ADEMOVIC, A., NUHANOVIC, A., “Correlation between Air Temperature and Electricity Demand by Linear Regression and Wavelet Coherence Approach: UK, Slovakia and Bosnia and Herzegovina Case Study”, Archives of Electrical Engineering, 62 (4), 521-532, 2013.
- BERTHOME J. P., « Marel : un réseau automatisé de veille pour l'environnement littoral », Equinoxe (Nantes), n° 47-48, pp. 34-35, 1994.
- BLACKMAN R.B., and TUKEY J.W., “The measurement of power spectra from the point of view of communication engineering”, Dover Publications, 190 pp, 1958.
- BURCHARD H., BAUMERT H., The formation of estuarine turbidity maxima due to density effects in the salt wedge. A hydrodynamic process study”, J. Phys. Oceanogr. 28, 309-321, 1998.
- CHALLIS R. E. and KITNEY R. I., “Biomedical signal processing (in four parts). Part 1 Time-domain methods. Medical & Biological Engineering & Computing”, 28, November 1991.
- CHANG G., DICKEY T. D., “Optical and physical variability on timescales from minutes to the seasonal cycle on the new England shelf: July 1996 to June 1997”, J. Geophys. Res. 106, 9435–9453, 2001.
- CHANT R. J., STONER A. W., “Particle trapping in a stratified flood-dominated estuary”, J. Marine Res. 59, 29-51, 2001.
- CHAVEZ F. P., PENNINGTON J. T., HERLIEN R., JANNASCH H., THURMOND G., and FRIEDERICH G. E., “Moored and drifters for real-time interdisciplinary oceanography”, J. Atmos. Ocean. Technol. 14, 1199–1211, 1997.
- CHEN, X., WU, Z., HUANG, N.E., “The time-dependent intrinsic correlation based on the empirical mode decomposition”, Adv. Adapt. Data Anal. 2, 233–265, 2010.
- DÄTIG, M., SCHLURMANN, T., “Performance and limitations of the Hilbert–Huang transformation (hht) with an application to irregular water waves”, Ocean Eng. 31 (14), 1783–1834, 2004.
- DICKEY T. D., “The emergence of concurrent high resolution physical and bio-optical measurements in the upper ocean and their applications”, Rev. Geophys. 29, 383–413, 1991.
- DICKEY T.D., DOUGLASS R.H., MANOV D., and BOGUCKI D., “An experiment in duplex communication with a multi-variable moored system in coastal waters”, J. Atmos. Ocean. Tech, 10, 637-644, 1993.
- DICKEY D. A., FULLER W. A., “Distribution of the estimates for autoregressive time series with a unit root”, J. Am. Stat. Assoc., 74, 427-431, June 1979.
- DIKS C., “Nonlinear Time Series Analysis: Methods and Applications”, World Scientific Press, 180 pp, 1999.
- DUR G., SCHMITT F., SOUISSI S., “Analysis of high frequency temperature time series in the Seine estuary from the Marel autonomous monitoring buoy”, Hydrobiologia, 588, 59-68, 2007.
- EMERY W. J., and THOMSON R. E., “Data Analysis Methods in Physical Oceanography”, 2014.
- EZER, T., ATKINSON, L. P., CORLETT, W. B., BLANCO, J. L., “Gulf Stream’s induced sea level rise and variability along the U.S. mid-Atlantic coast”, J. Geophys. Res. Oceans, 118, 685–697, 2013.
- FLANDRIN, P., GONÇALVÈS, P., “Empirical mode decompositions as data-driven wavelet-like expansions”, Int. J. Wavelets, Multires. Info. Proc. 2 (4), 477-496, 2004.
- FLANDRIN, P., RILLING, G., GONÇALVÈS, P., “Empirical mode decomposition as a filter bank”, IEEE Signal Proc. Lett. 11 (2), 112–114, 2004.

- FOREMAN M., “Manual for tidal heights analysis and prediction”, Technical Report Pacific Marine Report 77-10, Institute of Ocean Sciences, Patricia Bay Victoria, BC, Canada, 1977, revised 1996.
- GONI M. A., CATHEY M. W., KIM Y. H., and VOULGARIS G., “Fluxes and sources of suspended organic matter in an estuarine turbidity maximum region during low discharge conditions”, *Estuarine Coastal Shelf Sci.*, 63, 683–700, 2005.
- HUANG, N., SHEN, Z., LONG, S.R., WU, M.C., SHIH, H.H., ZHENG, Q., YEN, N., TUNG, C.C., LIU, H.H., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”, *Proc. R. Soc. London, Ser. A* 454 (1971), 903–995, 1998.
- HUANG, N., SHEN, Z., LONG, S.R., “A new view of nonlinear water waves: the Hilbert spectrum”, *Annu. Rev. Fluid Mech.* 31 (1), 417–457, 1999.
- HUANG, Y., “Arbitrary-order Hilbert Spectral Analysis: Definition and Application to Fully Developed Turbulence and Environmental Time Series”, (Ph.D. thesis) Université des Sciences et Technologies de Lille — Lille 1, France & Shanghai University, China, 2009.
- HUANG, Y., SCHMITT, F.G., “Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition”, *J. Mar. Syst.*, 130, 90–100, 2014.
- HUANG, Y., SCHMITT, F., LU, Z., LIU, Y., “An amplitude–frequency study of turbulent scaling intermittency using Hilbert spectral analysis”, *Europhys. Lett.* 84, 40010, 2008.
- IBANEZ F., CONVERSI A., “Prediction of missing values and detection of ‘exceptional events’ in a chronological planktonic series: a single algorithm”, *Ecol. Modelling*, 154, 9-23, 2002.
- KANTZ H., and SCHREIBER T., “Nonlinear Time Series Analysis”, Cambridge University Press, 304 pp, 1997.
- KENDALL M. G., “Oscillatory movements in English agriculture (with discussion)”, *Journal of the Royal Statistical Society*, 106, 91-124, 1943.
- LAU K. M., WENG H., “Climate Signal Detection Using Wavelet Transform: How to make a Time Series Sing”, *Bulletin of the American Meteorological Society*, 76, 2391-2402, 1995.
- LOMB, N. R., “Least-Squares Frequency Analysis of Unevenly Spaced Data”, *Astrophysical and Space Science*, 39, 447-462, 1976.
- MACKINNON J., “Numerical distribution function for unit root and cointegration tests,” *J. Appl. Econ.*, 11, 601–618, 1996.
- MOGHTADERI, A., BORGNAT, P., FLANDRIN, P., “Trend filtering: empirical mode decompositions versus 11 and Hodrick–Prescott”, *Adv. Adapt. Data Anal.* 3 (01n02), 41–61, 2011.
- MONIN A. S., “Theoretical Geophysical Fluid Dynamics”, 399 pp., Kluwer Acad. Publ., Dordrecht, 1990.
- NELSON C. and PLOSSER C., “Trends and random walks in macroeconomics time series: some evidence and implications”, *J. Monet. Econ.*, 10, 130–162, 1982.
- NG S. and PERRON P., “Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag”, *Journal of the American Statistical Association*, 90, 268-281, 1995.
- PANOFKY H. A. and BRIER, G. W., “Some applications of statistics to meteorology”, The Pennsylvania State University, College of Earth and Mineral Sciences, University Park, 224 pp, 1968.
- PAPADIMITRIOU, S., SUN, J., YU, P.S., “Local correlation tracking in time series”, *Proc. Sixth Int. Conf. Data Mining* 456–465, 2006.
- PAPARELLA F., “Filling gaps in chaotic time series”, *Phys. Lett. A*, 346, 47-53, 2005.
- PAWLOWICZ R., BEARDSLEY B., and LENTZ S., “Classical tidal harmonic analysis including error estimates in MATLAB using T\_TIDE”, *Computers and Geosciences* 28, 929-937, 2002.
- PRESS W.H., TEUKOLSKY S.A., VETTERLING W.T., “Numerical Recipes in Fortran 77. Cambridge University Press, Cambridge, 1992.

- RODO, X., RODRIGUEZ-ARIAS, M.A., “A new method to detect transitory signatures and local time/space variability structures in the climate system: the scale-dependent correlation analysis”, *Clim. Dyn.* 27, 441–458, 2006.
- SCARGLE J. D., “Studies in Astronomical Time Series Analysis. I. Modeling Random Processes in the Time Domain”, *The Astrophysical Journal Supplement Series* 45, 1–71, 1981.
- SCARGLE J. D., “Studies in Astronomical Time Series Analysis. II. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data”, *The Astrophysical Journal* 263, 835–853, 1982.
- SCARGLE J. D., “Studies in Astronomical Time Series Analysis. III. Fourier Transforms, Autocorrelation Functions, and Cross-Correlation Functions of Unevenly Spaced Data”, 1989.
- SCHMITT F. G., DUR G., SOUISSI S., BRIZARD ZONGO S., “Statistical properties of turbidity, oxygen and pH fluctuations in the Seine river estuary (France)”, *Physica A: Statistical Mechanics and its Applications*, 387, 6613–6623, 2008.
- SCHMITT, F.G., HUANG, Y., LU, Z., LIU, Y., FERNANDEZ, N., “Analysis of velocity fluctuations and their intermittency properties in the surf zone using empirical mode decomposition”, *J. Mar. Syst.* 77, 473–481, 2009.
- SCHULZ M., STATTEGGER K., “SPECTRUM: Spectral Analysis of Unevenly Spaced Paleoclimatic Time Series”, *Computers & Geosciences* 23, 929–945, 1998.
- SCHUREMAN P. , “Manual of Harmonic Analysis and Prediction of Tides”, *Special Publication* 98, Coast and Geodetic Survey, U.S. Dept. of Commerce, Washington, D.C., 317 pages, 1958. (Revised 1940 Edition with corrections; first published in 1924).
- SCHWERT W., “Test for Unit Roots: A Monte Carlo Investigation”, *Journal of Business and Economic Statistics*, 7, 147–159, 1989.
- THOMSON SIR WILLIAM (LORD KELVIN), “Report of the Committee for the purpose of promoting the extension, improvement and harmonic analysis of Tidal Observations”, In: *Report of the 38th Meeting of the BAAS*, 1868, 489–510, John Murray, London, 1869.
- THOMSON SIR WILLIAM (LORD KELVIN), “The tide gauge, tidal harmonic analyzer and tide predictor”, *Proceedings Institute of Civil Engineering*, London, 65, 4–74, 1881.
- TONG H., “Nonlinear Time Series Analysis”, Oxford University Press, 564 pp. Windrows, B., and S. D. Stearns, 1985: *Adaptive Signal Processing*. Prentice Hall, 474 pp, 1990.
- TORRENCE, C., COMPO, G., “A practical guide to wavelet analysis”, *Bulletin of the American Meteorological Society* 79 (1), 61–78, 1998.
- WELCH P.D., “The Use of Fast Fourier Transform for the Estimation of Powerspectra: A Method Based on Time Averaging over Short, Modified Periodograms”, *IEEE Trans. Audio Electroacoustics* AU-15, 70–73, 1967.
- WOERTHER P., « Marel, Mesures Automatisées en Réseau pour l'Environnement Littoral », *L'eau, l'Industrie, les Nuisances*, 217, 67–71, 1998.
- WOLD H., “A study in the analysis of stationary time series”, Uppsala: Almqvist & Wiksell, 1938.
- WU, Z., HUANG, N.E., “A study of the characteristics of white noise using the empirical mode decomposition method”, *Proc. R. Soc. Lond. Ser. A* 460, 1597–1611, 2004.
- WU, Z., HUANG, N.E., LONG, S.R., PENG, C., “On the trend, detrending, and variability of nonlinear and non-stationary time series”, *PNAS* 104 (38), 14889, 2007.
- YIN, L., QIAO, F., ZHENG, Q., “Coastal-trapped waves in the East China Sea observed by a mooring array in winter 2006”, *J. Phys. Oceanogr.*, 44, 576–590, 2014.
- YULE G. U., “Why do we sometimes get nonsense correlations between time series? - A study in sampling and the nature of time series (with discussion)”, *Journal of the Royal Statistical Society*, 89, 1–63, 1926.
- ZONGO S., SCHMITT F., “Scaling properties of pH fluctuations in coastal waters of the English Channel: pH as a turbulent active scalar”, *Nonlinear Processes in Geophysics*, 18, 82, 2011.